AD-A081 119 AIR FORCE ACADEMY CO F/G 5/9
THE H2 EXPERIMENT: A COMPARISON OF HOMOGENEOUS AND HETEROGENEOU--ETC(U)
DEC 79 S B THOMPSON, C R MITCHELL, R C COFFIN
UNCLASSIFIED USAFA-TR-79-9 NL

1 OF 1
AD
A081119

END
DATE
FILMED
3-80
DDC

USAFA TR 79-9

THE H2 EXPERIMENT: A COMPARISON OF HOMOGENEOUS AND HETEROGENEOUS APTITUDE SECTIONING IN CORE MATHEMATICS

SAMUEL B. THOMPSON, MAJOR, USAF
CHARLES R. MITCHELL, LT COLONEL, USAF
ROBERT C. COFFIN, CAPTAIN, USAF
MATTHEW J. HASSETT, DISTINGUISHED VISITING PROFESSOR

DECEMBER 1979

FINAL REPORT

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED

DEAN OF THE FACULTY
UNITED STATES AIR FORCE ACADEMY
COLORADO 80840

Technical Review by Colonel Seegmiller
Department of Chemistry and Biological Sciences
USAF Academy, Colorado 80840

Technical Review by Major Smith
Department of Mathematical Sciences
USAF Academy, Colorado 80840

Editorial Review by Captain Anderson
Department of English
USAF Academy, Colorado 80840

This research report is presented as a competent treatment of the subject, worthy of publication. The United States Air Force Academy vouches for the quality of the research, without necessarily endorsing the opinions and conclusions of the author.

This report has been cleared for open publication and/or public release by the appropriate Office of Information in accordance with AFR 190-17 and DODD 5230.9. This is no objection to unlimited distribution of this report to the public at large, or by DDC to the National Technical Information Service.

This research report has been reviewed and is approved for publication.

John P. Wittry

JOHN P. WITTRY, Colonel, USAF
Vice Dean of the Faculty

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER (14) USAFA-TR-79-9 ✓	2. GOVT ACCESSION NO	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) The H2 Experiment: A Comparison of Homogeneous and Heterogeneous Aptitude Sectioning in Core Mathematics.		5. TYPE OF REPORT & PERIOD COVERED (9) Final Report,
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Samuel B. Thompson, Maj, USAF; Charles R. Mitchell, LtCol, USAF; Robert C. Coffin, Capt, USAF; Matthew J. Hassett, Distinguished Visiting Prof		8. CONTRACT OR GRANT NUMBER(s)
9. PERFORMING ORGANIZATION NAME AND ADDRESS Dept of Mathematical Sciences, United States Air Force Academy, CO 80840		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS USAFA/DFMS, USAF Academy, CO 80840		12. REPORT DATE (11) December 1979
		13. NUMBER OF PAGES 29
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) (12) 41		15. SECURITY CLASS (of this report) Unclassified
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Homogeneous, Heterogeneous, Aptitude Grouping, Experimental Comparison, College Calculus, Sectioning Effects

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The H2 experiment was a comprehensive large scale study of the relative merits of homogeneous and heterogeneous sectioning modes for the purpose of determining sectioning policy in core mathematics courses. In Calculus II, 429 cadets were homogeneously sectioned and 428 were heterogeneously sectioned. Thirteen and twelve instructors were respectively assigned to the homogeneous and heterogeneous sections. Cadets and instructors were assigned via stratified random procedures and extraneous variables which could affect measured outcomes were

(continued on reverse)

DD FORM 1 JAN 73 1473

011550

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20. Abstract (continued)

either balanced or randomized in the design. Measured outcomes included math achievement, fail rate, cadet outside study time, EI workload and instructor opinion. Overall, no statistically significant mean differences were attributable to sectioning mode.

THE H2 EXPERIMENT: A COMPARISON OF HOMOGENEOUS AND HETEROGENEOUS APTITUDE SECTIONING IN CORE MATHEMATICS

Samuel B. Thompson, Major, USAF
Charles R. Mitchell, Lt Colonel, USAF
Robert C. Coffin, Captain, USAF
Matthew J. Hassett, Distinguished Visiting Professor

October 1979

ABSTRACT

The H2 experiment was a comprehensive large scale study of the relative merits of homogeneous and heterogeneous sectioning modes for the purpose of determining sectioning policy in core mathematics courses. In Calculus II, 429 cadets were homogeneously sectioned and 428 were heterogeneously sectioned. Thirteen and twelve instructors were respectively assigned to the homogeneous and heterogeneous sections. Cadets and instructors were assigned via stratified random procedures and extraneous variables which could affect measured outcomes were either balanced or randomized in the design. Measured outcomes included math achievement, fail rate, cadet outside study time, EI workload and instructor opinion. Overall, no statistically significant mean differences were attributable to sectioning mode.

ACKNOWLEDGEMENTS

The H2 experiment was a cooperative effort of many people within the Department of Mathematical Sciences. LtCol Jon Epperson, Capt Ray Cardaronella, and Capt Russell Webster performed key functions in the design and conduct of the experiment. LtCol Jay Sherman chaired the experimental group and drafted the experimental plan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

Since the founding of the US Air Force Academy, cadets have been sectioned homogeneously according to aptitude in core mathematics courses. In this procedure, a predictor variable which is related to aptitude is used to rank order cadets prior to section assignment. Examples of predictor variables are academic composite score, placement test score or score in the previous math course. The top n cadets in the rank order are assigned to the "A" class section, the second n cadets to the "B" class section and so on until all cadets are assigned. This practice was inherited from the US Military Academy and dates from 1818.[1]

Homogeneous sectioning is intended to accommodate a higher degree of "instructional tailoring" to the specific needs of individual cadets than is possible under other group-oriented instructional strategies. When aptitude variance in the class section is minimized, a single pace and level of presentation are presumed to be more appropriate for each cadet. Historically, homogeneous sectioning has also been used as a behavioral incentive for improved performance. When homogeneous sectioning is used for this purpose, prestige is conferred on the highest sections through introduction of advanced material, instruction from high-ranking faculty or some other means. Assignment to a low section is not prestigious and cadets are resectioned monthly or after every graded review. The practice of resectioning during the course was abandoned several years ago in core mathematics courses at the USAFA.

Homogeneous aptitude sectioning is well publicized as a positive feature of the USAFA academic system. The hypothesis that homogeneous sectioning enriches the learning environment more than does random (heterogeneous) sectioning has been accepted as an article of faith, but at least in core mathematics, this hypothesis has not been empirically verified. Some verifications have been attempted, but these have not adequately tested the hypothesis under sound experimental conditions.

---

[1] W.E. Simons, Liberal Education in the Service Academies, Teachers College, Columbia University, 1965, p 41.

Cadets were heterogeneously sectioned in Math 131 during the Fall 1977 and Fall 1978 semesters as an incidental result of an educational experiment on teaching strategies. For most math faculty, these two episodes provided the first opportunity to experience an alternative sectioning scheme in a large core mathematics course. Heterogeneous sectioning made a favorable impression on several faculty members. In their opinions, heterogeneous sectioning improved cadet interaction in the classroom, provided mathematically deficient cadets with successful cadet role models, and eliminated the frustration of whole class sections of disinterested and mathematically deficient cadets. As a result, the Math 132 course director requested permission to heterogeneously section cadets in that course in the Spring 1979 semester. To obtain facts on the relative merits of homogeneous and heterogeneous sectioning, the Acting DFMS Head directed that an experimental comparison be conducted in Math 132, the second core mathematics course, during the Spring 1979 semester. That comparison is described and results reported in this document.

In Chapter II, the experimental design and methodology are outlined. In Chapter III, analyses and results are discussed. Finally, in Chapter IV, conclusions are drawn and outcomes summarized. Unless otherwise stated, results of statistical tests of hypotheses are reported with a confidence coefficient of 0.95.

Chapter II

DESIGN AND METHODOLOGY

In this chapter the experimental design used to compare homogeneous and heterogeneous modes of sectioning is described. Operational methods used in conducting the experiment are also explained. The experiment was conducted in Math 132, Calculus II, a three-unit every day course in which all cadets who successfully completed Math 131 in the Fall 1978 semester were enrolled. The experiment was designed to satisfy several criteria. One criterion was to control extraneous factors such as instructor experience and cadet aptitude, which might mask any effects of treatment. A second criterion was to obtain sufficient statistical power in the experimental design to detect any differences in educational or administrative outcomes caused by treatment which might be of practical importance. A third criterion was to measure a sufficiently broad range of outcomes such as Math 132 achievement, cadet study time, EI workload and fail-rate to afford a comprehensive perspective of treatment effects. Measured outcomes are called *dependent variables* throughout the report. Mode of sectioning is called *treatment* with homogeneous sectioning constituting one level of treatment and heterogeneous sectioning constituting another level. Factors such as cadet aptitude, instructor experience and treatment which were controlled and analyzed for their effects on measured outcomes are called *independent variables*. The initial plan of the experiment is included as Appendix A.

Cadet Assignment to Treatment

To understand the assignment of cadets to treatment, it is first necessary to understand the measure of "aptitude" used in the study. This is an important point because equality of mean "aptitudes" was the major criterion in assignment of cadets to cells in the experimental design. Cadet "aptitude" is really a convenient misnomer in this report. The measure of aptitude used throughout this study was Math 131 percent score although it is known that this score is affected by motivation and other factors than aptitude. Math 131 percent score was the strongest

available predictor of Math 132 percent score having a correlation in the total experimental population of 0.76.

Using the master scheduling algorithm maintained by the Department of Curriculum and Scheduling, 856 cadets were assigned to Math 132 in class period 1, 2, 3, 4, 5, 6, or 7. The mean aptitude of the cadets assigned in each period of the day was computed and the means were tested in a one-way analysis of variance (ANOVA) to determine whether they differed significantly from period-to-period. They did not (see Tables B1 and B2, Appendix B).

Cadets in periods 2 and 6 were then sectioned homogeneously on aptitude as described previously. Cadets in periods 3 and 4 were sectioned heterogeneously according to the following procedure: (a) cadets were initially rank ordered on aptitude as in homogeneous sections; (b) the first cadet in the rank order was assigned to section A, the second to B and so on until one cadet was assigned to each of the k sections in the period; (c) the next k cadets were assigned to sections in reverse order so that the 2kth cadet was in the A section. Steps (b) and (c) were repeated until all cadets were assigned. In each of periods 1, 5, and 7, half of the cadets were homogeneously sectioned and half heterogeneously sectioned. The procedure here was to initially rank order all cadets assigned to the period on aptitude. Odd numbered cadets in the rank ordering were homogeneously sectioned. Even numbered cadets were heterogeneously sectioned as described in (a),(b), and (c) above.

The reason for the particular sectioning scheme was the hypothesis that period of the day could itself affect cadet performance. According to this hypothesis, performance in periods 2, 3, 4 and 6 would be similar but performance in periods 1, 5 and 7 would be systematically different; lower in period 1 because cadets are not fully awake, lower in period 5 because of drowsiness from lunch, and higher in period 7 because on-season athletes are usually not assigned in that period. The hypothesis was tested after the experiment in a one-way analysis of covariance with aptitude as covariate and was rejected. No significant differences in Math 132 percent score could be attributed to the effect of period of the day (see Tables B1 and B3, Appendix B).

Instructor Assignment to Treatment

The 25 Math 132 instructors were categorized as to experience level and assigned to treatment in such a way as to balance instructor experience in both treatments. Results of the assignment scheme are shown in Table 1. Instructors did not cross treatments.

Table 1. Instructor Experience Breakdown

Experience Level	Treatment[a] Homo	Hetero	Prior Teaching Experience at AFA
1	4	3	None
2	4	3(1)	≤ 2 years
3	4(1)	5	> 2 years

[a]Numbers in parentheses indicate instructors who participated in the experiment but left prior to administration of the post-treatment survey. These instructors are not included in the instructor opinion analysis but their students are included in the achievement analysis.

In addition to experience level, instructor forces were also balanced on treatment preference, number of rated instructors and number of split-loaded instructors.

Methodology

Instructors with homogeneous sections were advised to provide ample drill for low sections and enrichment material for high sections. Instructors with heterogeneous sections were advised to use stronger students to help weaker students in their classes. As is customary in core math courses, instructors were given wide personal latitude; no attempt was made to insure that the given advice was acted upon.

To minimize reactive effects which might contaminate measurements, a low profile was maintained for the experiment. Just prior to the experiment, instructors were told that an experimental study on effects of homogeneous and heterogeneous sectioning would be conducted in Math 132. Thereafter, no further mention was made of the experiment

until after the conclusion of Math 132. All comparative data were acquired from sources normally available in core math courses. That is, no measurements unique to the experiment which might have heightened awareness of the experiment and perturbed outcomes were made during the experiment. Data were accumulated by an independent investigator outside Math 132. No comparative analyses were performed until all data were acquired. Without prior knowledge that an experiment was underway, one would not have detected that fact from day-to-day observations in Math 132.

Several dependent variables were analyzed in the study to minimize the possibility that effects of educational or administrative importance might escape detection. Dependent variables included math achievement, fail rate, cadet out-of-class study time, EI workload and instructor opinion.

Chapter III

ANALYSES AND RESULTS

In this chapter, the measurement and analysis procedures and the results for each dependent variable are described separately.

Math Achievement

Two dependent variables for math achievement were employed in the study: (1) total Math 132 percent score which was the sum of 32 quiz scores (33% of total), 3 graded review scores (35%), one final exam (29%), and 100 instructor prerogative points (3%); and (2) Math 132 final exam score. The first variable, while more comprehensive, was subject to contamination by factors other than achievement. For example, the allocation of instructor prerogative points may have been based on other criteria than achievement. The second variable was invulnerable to such contaminants because final exams were standardized in content and administration and were team-graded. It turned out that measurements on either variable led to precisely the same results. Therefore, only the results on total Math 132 percent score are included in this report.

Math 132 percent score was treated as the dependent variable in two modes of statistical analysis, analysis of variance (ANOVA) and regression analysis.

Analysis of Variance (ANOVA)

Analysis of variance is a statistical technique in which the total variation in values of a dependent variable is partitioned, and components of the variation are ascribed to each independent variable under test for significance on the dependent variable[2]. An ANOVA was conducted on 857 cases in a factorial design with three independent variables: treatment, cadet aptitude and instructor experience. There were two levels of treatment, *homogeneous* sectioning and *heterogeneous* sectioning. There were three levels of aptitude; *low* corresponding to Math 131 percent scores <72.6, *medium*, corresponding to scores ≥72.6 but <80.5 and *high*

[2]For a detailed discussion of ANOVA see G.V. Glass and J.C. Stanley, Statistical Methods in Education and Psychology, Prentice-Hall, 1970, pp 338-380.

corresponding to scores ≥80.5. These particular divisions were chosen solely to obtain nearly equal cell frequencies in the ANOVA. There were two levels of instructor experience; *inexperienced* which was synonymous with level 1 in Table 1 and *experienced* which was an aggregate of levels 2 and 3 in Table 1. Levels 2 and 3 were combined solely to obtain nearly equal cell frequencies in the ANOVA. In the ANOVA, instructor experience level was shown to have no significant effect on cadet performance in Math 132 (see Tables B4 and B5, Appendix B). That is, instructors with virtually no prior classroom teaching experience at the USAFA produced indistinguishable performance results from experienced faculty. Similarly no significant interaction of instructor experience with treatment or with cadet aptitude level was detected (see Tables B4 and B5, Appendix B). Furthermore, no three-way interaction of the three factors was detected (see Tables B4 and B5, Appendix B).

As a result of the ANOVA described in the previous paragraph, the instructor experience factor was collapsed in the design and effects of treatment and cadet aptitude were re-analyzed in a two-way factorial ANOVA to increase statistical power. Cell means, standard deviations and cell frequencies are shown in Table 2. Detailed ANOVA results are shown in Table 3.

Table 2. Cell Statistics of Math 132 Percent Scores

Treatment	Cadet Aptitude Level	Mean	Std Dev	Freq
Homogeneous Sectioning		79.1	10.8	
	Low	69.2	9.1	141
	Medium	79.8	7.3	147
	High	88.3	5.7	141
Heterogeneous Sectioning		78.6	10.3	
	Low	69.6	8.8	139
	Medium	78.6	7.5	146
	High	87.3	5.7	143

Table 3. Two Factor ANOVA Results on Math 132 Percent Score

Source	df	MS	F
Treatment (T)	1	78.3	1.41
Cadet Aptitude (A)	2	23896.8	429.04*
T x A	2	54.5	.98
Residual	851	55.7	

*p<.001

Salient results from Tables 2 and 3 are as follows:

1. Mean Math 132 percent scores of the homogeneously and heterogeneously sectioned cadets were not statistically different at even the .8 confidence level.

2. Cadet aptitude, as measured by Math 131 percent score, had a strong effect on Math 132 performance in both sectioning schemes.

3. Aptitude-treatment interaction was not significant at even the .8 confidence level. This is an important result since it was widely hypothesized that heterogeneous sectioning would "help" weak students but would "hurt" strong students. An aptitude treatment interaction diagram is illustrated in Figure 1. A slight effect in support of the hypothesis is observable due to the high magnification of the vertical scale. However, this effect is weak; the probability that the effect is due to chance is .38, about two in five.

Certain other results were obtained in ancillary analyses of variance:

4. Period of the day was not a significant factor affecting Math 132 performance in either sectioning scheme.

5. It was hypothesized that the performance of a class section might be enhanced by labeling it "A" and depressed by labeling it "F." A specific test for effect of labeling was built into the experimental design. The test produced a null result. Labeling sections as "A," "B," or "C" as contrasted with "D," "E," or "F" had no detectable effect on Math 132 performance.

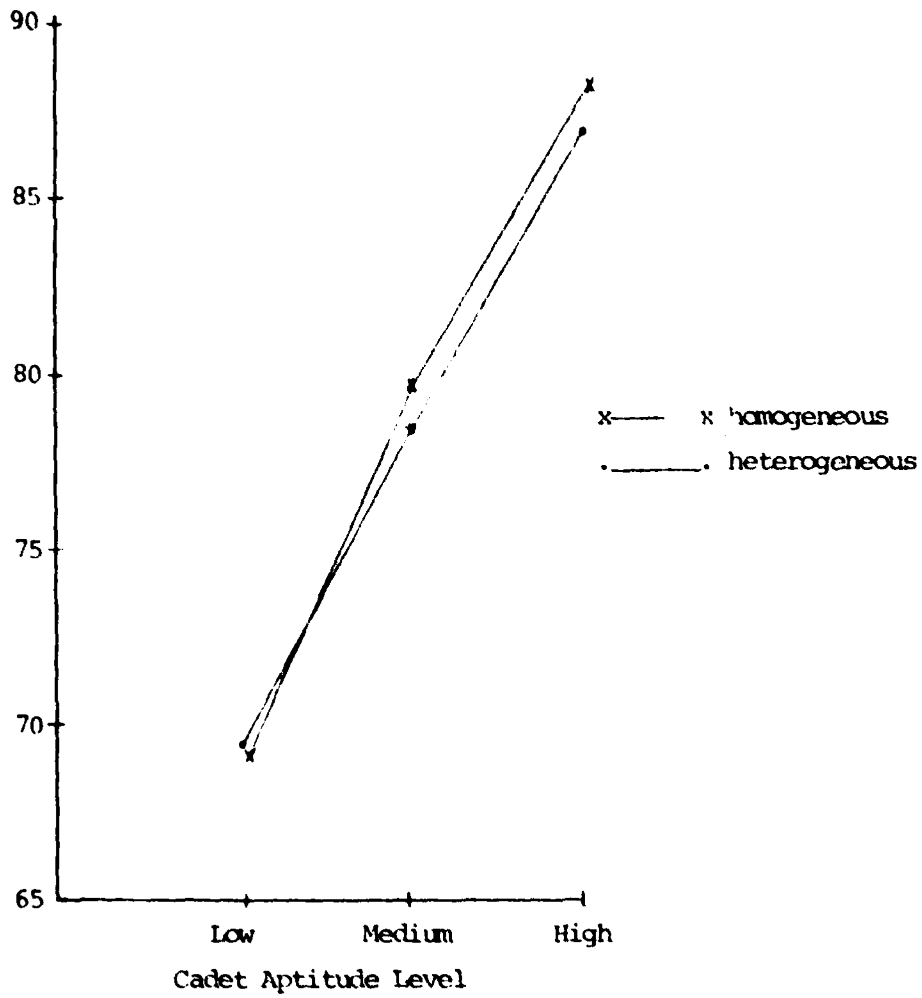

Figure 1. Aptitude-Treatment Interaction Diagram

Regression Analysis

To verify the ANOVA results, Math 132 achievement data were re-analyzed via multiple regression procedures. While the fundamental statistical assumptions underlying regression analysis and analysis of variance are the same, the problem is formulated in different ways and different computational techniques are employed. Regression analysis, therefore, provided a cross-check of the ANOVA results.

The following linear regression model was assumed:

$$M132' = B_0 + B_1T + B_2A + B_3 (A \times T)$$

where

M132' is the predicted Math 132 percent score

T is treatment and has value 0 for homogeneous sectioning and 1 for heterogeneous sectioning

A is Math 131 percent score and

A x T is the product of A and T

The model attempts to predict Math 132 scores from type of sectioning (treatment), Math 131 score (aptitude) and the aptitude-treatment interaction term (A x T). B_1, B_2, and B_3 are the respective regression coefficients. B_0 is a constant. In the regression analysis, B_0, B_1, B_2, and B_3 are chosen to give the "least squares" best fit to the observed Math 132 percent scores. Results of the regression analysis are shown in Table 4.

Table 4. Multiple Regression Results

Independent Variable	Regression Coefficient (B_i)	i	Percent Variance Explained
Treatment (T)	6.64	1	0.1
Aptitude (A)	1.10*	2	57.0
A x T	-.10	3	0.2
Constant	-4.80	0	

*$p < .001$

Using this model, 57.3% of the variance in Math 132 percent scores can be explained. However, 57% of the variance is explained by the aptitude variable. The probability that B_2 is significantly different from zero is >.999. On the other hand, neither B_1 nor B_3 differs significantly from zero. Thus aptitude is the only significant predictor of Math 132 performance. Graphically, this can be seen from the following analysis in which the two regression lines suggested by the above model are compared.

The above model reduces to

$$M132' = B_0 + B_2 A = -4.80 + 1.10\ M131$$

for homogeneous sectioning and

$$M132' = B_0 + B_1 + (B_2 + B_3)A = 1.84 + 1.00\ M131$$

for heterogeneous sectioning. These two lines are graphed in Figure 2. Low aptitude cadets did slightly better under heterogeneous sectioning and high aptitude cadets under homogeneous sectioning but this effect was very weak and very likely due to chance. Conclusions drawn from Figure 2 are the same as those drawn from Figure 1.

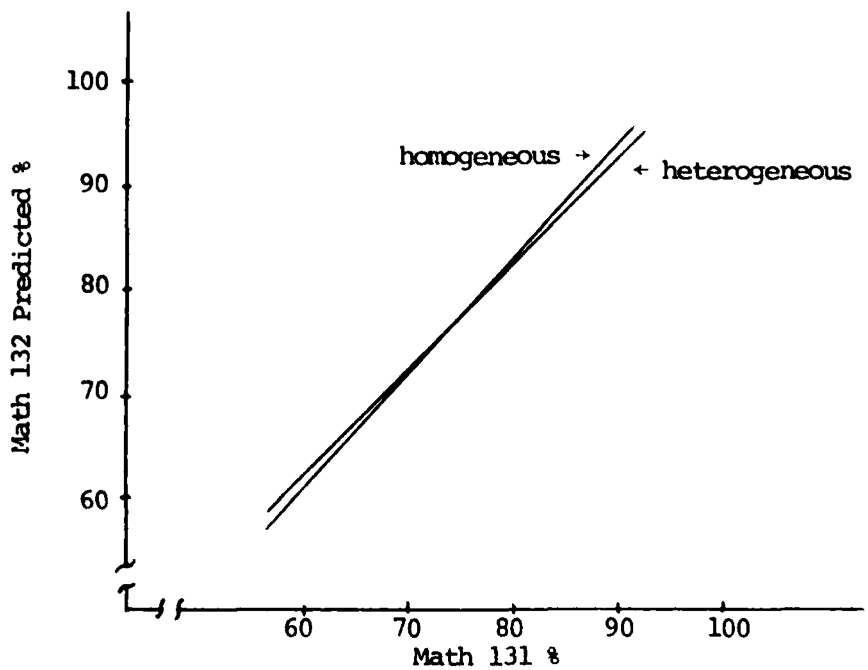

Figure 2. Regression Lines for Homogeneous and Heterogeneous Sectioning

Another linear regression model incorporating period of the day and instructor experience level as additional independent variables was investigated. Neither variable contributed significantly to the explained variance.

All strengths of effects determined in the regression analysis were in complete agreement with results of the analysis of variance.

Fail Rate

Fail rate is defined as the percent of students who took the final exam that received an F course grade. The Math 132 fail rate in homogeneous sections was 5.1%. In heterogeneous sections the fail rate was 4.7%. For all practical purposes, the fail rates were the same in both treatments.

Cadet Study Time

Self-reported out-of-class study times of homogeneously and heterogeneously sectioned cadets were compared to detect any difference in level of cadet effort attributable to treatment. The typical study time measurement method in core math courses is to request anonymous reporting of preparation time on a form circulated in the class section. Such forms are distributed without prior announcement to randomly selected sections on randomly selected lessons. Many possible sources of invalidity can be identified in this measurement method. For example, the operational definition of out-of-class study time is not standardized among cadets and they may report based on different definitions. Other measurement deficiencies cloud results of the treatment comparison. For example, the sampling scheme was not stratified on the all-important aptitude variable; the relative proportions of high and low aptitude cadets that were sampled differed drastically between treatments.

Homogeneous sections were sampled 84 times and heterogeneous sections were sampled 82 times. Based on these data, the mean study time per cadet per lesson was 54.5 and 53.7 minutes respectively in the homogeneous and heterogeneous treatments. For all practical purposes, there was no difference in mean study time per lesson.

Extra Instruction

The total time expended in extra instruction (EI) was compared between treatments as one indicator of instructor workload. Upon completion of an EI session each instructor gave each cadet an EI slip on which to report number of minutes of EI received. Each cadet was directed to deposit the slip in a designated container in the math department office. Total EI time was computed from the slips. While several possible sources of invalidity in this measurement method are identifiable, none of these contribute to systematic difference between treatments.

Total EI times reported in the homogeneous and heterogeneous treatments were respectively 260.4 and 259.2 hours. Strikingly, the difference was only 68 minutes in 260 hours. For all practical purposes, the same amount of EI was administered in both treatments.

Instructor Opinion

To assess instructor opinion regarding homogeneous and heterogeneous sectioning, a pretreatment survey was administered one month prior to the start of the experiment and a post-treatment survey was administered one month after the experiment's conclusion. Sample survey forms are included as Appendix C. For purposes of analyzing instructor opinion, instructors were categorized in the three experience levels of Table 1.

The pretreatment survey was used primarily to balance instructor forces between treatments. This survey was not administered to instructors at experience level 1. By the time of the post-treatment survey all instructors assigned heterogeneous sections in the experiment had acquired some experience with homogeneous sectioning. Unfortunately, some instructors assigned homogeneous sections in the experiment had not acquired any experience with heterogeneous sectioning.

Twenty-three instructors responded to the first nine items on the post-treatment survey according to the scale at the top of Table 5. Survey items and mean responses are also shown in Table 5. The strongest opinion on these items was that homogeneous sectioning provides better learning conditions for cadets of high math aptitude.

Table 5. Mean Responses to Post-Treatment Survey Items

Homogeneous Sectioning			Heterogeneous Sectioning	
-2	-1	0	+1	+2
Strong Opinion	Moderate Opinion	No Difference	Moderate Opinion	Strong Opinion

Item	Mean Response
Which provides better learning conditions for:	
1. Cadets of low math aptitude?	0.00
2. Cadets of intermediate math aptitude?	-0.04
3. Cadets of high math aptitude?	-1.22
4. The majority of cadets?	-0.13
5. In which type of sectioning is the instructor's job more difficult?	+0.35
Which type of sectioning produces:	
6. More interaction between cadets in class?	+0.52
7. The higher EI load?	-0.26
8. The lower overall instructor workload?	0.00
9. In which type of sectioning would you rather teach core math?	-0.17

As can be seen in Figures 1 and 2, achievement data do not provide compelling support for this opinion. Weaker opinions were that heterogeneous sectioning produces more interaction between cadets in class (item 6), that the instructor's job is more difficult in heterogeneous sections (item 5), and that the EI load is heavier in homogeneous sections (item 7). As has been seen, this last opinion is not supported by EI data. Responses different from zero on every item but item 3 were close enough to zero to allow substantial possibility that they were chance fluctuations.

As a group, level 2 instructors in both treatments tended to favor homogeneous sectioning in their survey responses. Since more level 2 instructors than level 1 or level 0 instructors completed the survey, mean responses may have been slightly biased toward homogeneous sectioning due to the disproportionately large number of experienced instructors surveyed. Level 1 instructors as a group tended to favor heterogeneous sectioning. Level 0 instructors were about evenly split.

Item 9 in Table 5 was investigated in more detail. Since all level 1 and 2 instructors were asked, *Which type of sectioning would you prefer for Math 132: (a) homogeneous, (b) undecided, or (c) heterogeneous?* on the pretreatment survey, it was possible to evaluate changes in preference at these experience levels. Results are shown in Figure 3.

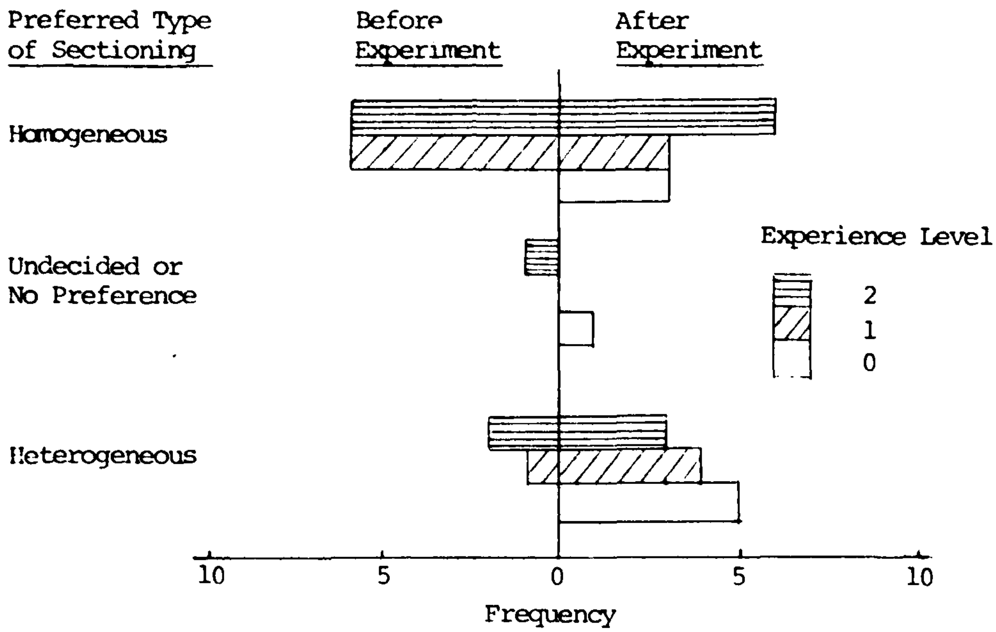

Figure 3. Changes in Sectioning Preference

Sectioning preference of level 2 instructors was not changed by the experiment except that one undecided instructor before opted for heterogeneous sectioning after the experiment. Three level 1 instructors changed their preference from homogeneous to heterogeneous sectioning. Three level 0 instructors opted for homogeneous sectioning and four for heterogeneous sectioning. The overall effect of the experiment was to swing preference toward heterogeneous sectioning.

A final item on the post treatment questionnaire was: *Which policy would you recommend to the Department Head for core math courses: (a) use heterogeneous sectioning only, (b) use homogeneous sectioning, (c) let individual course directors decide, or (d) other (explain)?*

The response pattern is represented in Figure 4.

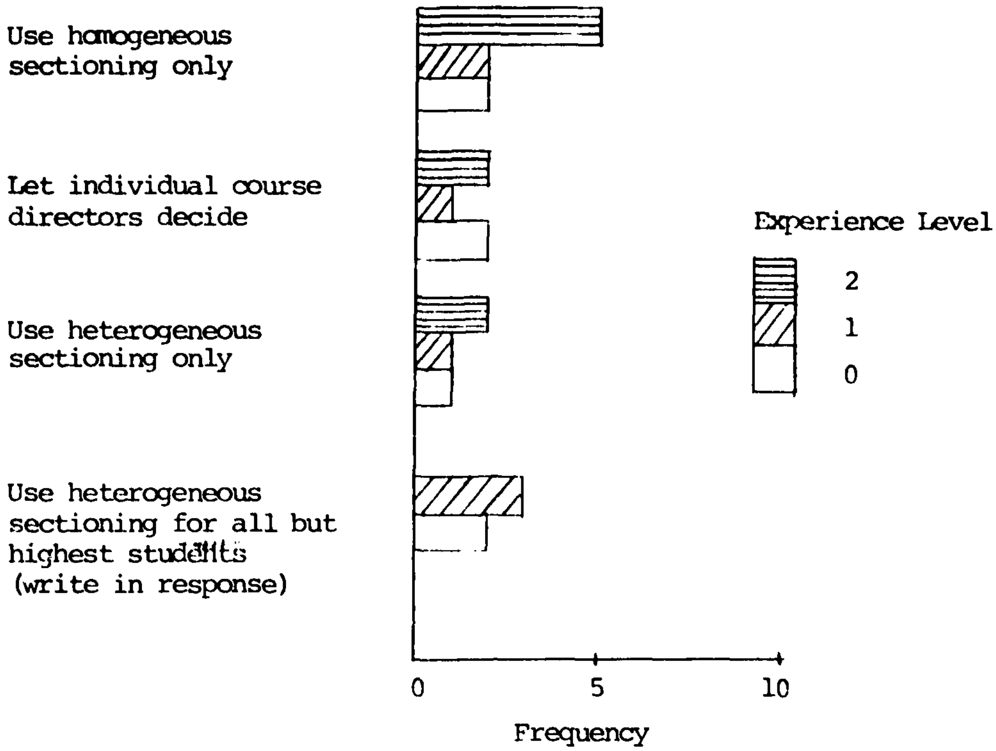

Figure 4. Core Math Sectioning Policy Recommendations

Level 2 instructors continued their trend of preference for homogeneous sectioning on this item. Five level 1 and level 0 instructors chose "other" and recommended heterogeneous sectioning for all but the highest aptitude students. Therefore, for the majority of cadets in core mathematics, level 1 and level 0 instructors indicated a preference for heterogeneous sectioning. Overall, the number of recommendations for each type of sectioning are equal for all but the highest aptitude students.

Chapter IV

DISCUSSION AND CONCLUSIONS

In this chapter, the various measured outcomes of the experiment are interwoven to afford a comprehensive perspective of the effects of homogeneous and heterogeneous sectioning.

The primary result of the experiment is that no significant differences attributable to homogeneous or heterogeneous sectioning were observed on any dependent variables measured on interval scales. Often this finding in educational experiments can be attributed to poor design or methodology. The effect of most flaws in balanced experiments is to enhance the probability of finding no significant differences. Flaws tend to mask actual treatment differences. While no experiment is absolutely flawless, flaws in this experiment were sufficiently minor to preclude masking of treatment differences of sufficient magnitude to be of any practical educational or administrative importance. In the H2 experiment, the conduct of large homogeneously and heterogeneously sectioned core math courses was well-simulated, confounding variables were controlled in a carefully balanced design, sample sizes were large, and the classic sources of internal invalidity[3] were minimized. For practical purposes, no difference in math achievement, fail rate, cadet study time, or EI workload can be attributed to the difference between homogeneous and heterogeneous sectioning.

One rather important result arising from the Math 132 scores is that math achievement seemed very insensitive to instructor experience. This result emerged in several ways. First, no achievement difference was attributable to the difference in instructor experience levels, i.e., after accounting for cadet aptitude differences, cadet achievement under experienced instructors was indistinguishable from achievement under inexperienced instructors. Second, no experience-treatment interaction was observed. Since the preponderance of instructor experience going into the experiment was with homogeneous sectioning, this experience

[3]For a detailed discussion of internal validity see D.T. Campbell, J.C. Stanley, Experimental and Quasi-experimental Designs for Research, Rand McNally, 1963, pp 1-17.

might have been expected to increment achievement under homogeneous sectioning. In particular, homogeneously sectioned high aptitude cadets might have been expected to score higher under senior faculty than their peers of the same aptitude who were assigned differently, but this was not the case. Third, if the instructor's performance on any given lesson is improved by experience with succeeding class sections, i.e., if the instructor does a better job with his second class of the day than with his first, there is no evidence in the results of this study that such improvement made any significant difference in cadet achievement. Individual instructors may enhance the achievement of some cadets (and depress the achievement of others) but there is no evidence that the individual instructor has any significant effect on math achievement of cadets when viewed macroscopically or as a group. If there are variables, other than cadet aptitude, which systematically affect the mean level of achievement in core math courses, t[illegible] variables lie outside the control of individual instructors. They may also lie outside the control of course directors and even outside the control of the Department of Mathematical Sciences. This is a very fruitful area for further study.

A very widely believed hypothesis was that homogeneous sectioning benefits the better math students. The strength of this belief is apparent in the response to item 3 of the instructor opinion survey (see Table 5) and in the frequency of the recommendation for heterogeneous sectioning for all but the highest students (see Figure 4). However, this hypothesis was not substantiated by the achievement data; the mean achievement of high aptitude students was slightly higher but not significantly higher under homogeneous sectioning. The difference was small enough to be attributable to measurement noise. Similarly, low aptitude students performed slightly better in heterogeneous sections but the difference was not significant.

As can be seen from Figures 3 and 4, the majority of senior faculty (level 2) preferred homogeneous sectioning before and after the experiment and recommended *homogeneous sectioning only* as departmental policy on core math sectioning. The majority of other faculty (levels

1 and 0) preferred heterogeneous sectioning after the experiment and recommended a department policy *other than* homogeneous sectioning only.

It may suffice to simply remark that "old heads" seemed to favor homogeneous sectioning distinctly more than less experienced faculty and let it go at that. But in a broad sense, the charter of the H2 study was not merely to make measurements and report results, it was to provide the most comprehensive perspective possible for sound educational policy regarding mode of sectioning. In light of the achievement results, why be concerned with faculty opinion or preference? The reason is that results of educational experiments do not generally enjoy the same credibility as results of experiments in the physical sciences. At the USAFA as in other institutions, educational policy decisions are often based on opinion, intuition, personal experience or some other basis than objective measurement. The senior faculty exert the most influence of any group in establishing, maintaining or altering educational policy within the institution and their opinions, even their preferences, could be crucial in the ultimate determination of sectioning policy. Therefore, if the collective opinion or preference of the senior faculty differs from the remainder of the faculty, the comprehensiveness of the perspective is enhanced not only by acknowledging this difference but also by inquiring into the reasons for it. Such an inquiry, conducted *a posteriori*, unfortunately requires conjecture. However, a further research effort to substitute facts for conjecture could be conducted.

One possibility is that the effect of experience in the USAFA is to eventually convince the faculty member that homogeneous sectioning is the superior mode here. Senior faculty could then be expected to hold fairly strong convictions favoring the homogeneous mode. Their mean response to item 4 in Table 5 might then also be expected to have a substantial negative value, say -1. But the mean response of level 2 instructors on item 4 was only -0.11 indicating that these officers did not necessarily think that homogeneous sectioning was better, *they just preferred it*. Perhaps we just prefer those situations with which we are most familiar.

The evidence from the comparison of pretreatment and post-treatment preference is that the net change in preference was in the direction of increased preference for heterogeneous sectioning. Perhaps if heterogeneous sectioning became more familiar, it would be more widely preferred. Perhaps if the two modes coexisted freely, a majority of faculty might even prefer heterogeneous sectioning within a few years.

Is there a relationship between sectioning preference and aptitude stratum of sections taught? Do senior faculty who prefer homogeneous sectioning usually teach high aptitude (A) sections? To explore the above possibility further during the writing of this report, a brief poll was taken of the seven DFMS faculty in the rank of Lieutenant Colonel and above. These were not the same officers as the level 2 instructors in the experiment. When asked item 9 in Table 5, three strongly preferred homogeneous sectioning, two moderately preferred it, and two moderately preferred heterogeneous sectioning. On the scale of Table 5, the mean response was -0.86 indicating, as in the initial H2 study, a definite preference for homogeneous sectioning. These same officers were also queried to determine the aptitude level of sections they were teaching or had most recently taught in large core math courses in which there were at least five sections per hour (A,B,C,D,E). Of the five senior faculty who had taught in such courses in recent semesters, three had A sections only, one had A and C sections and one had D and E sections. Of these five instructors, only the one who had D and E sections preferred heterogeneous sectioning. It seems quite likely that if the teaching assignments of senior faculty were distributed uniformly over the entire aptitude spectrum, their preference for homogeneous sectioning would diminish sharply.

Recommendation

The ensuing recommendation is based on the axiom that educational policy should only prescribe or preclude specific practices when such prescriptions or preclusions clearly promote the highest attainment of the goals of the USAFA program. Under this axiom, faculty are minimally

constrained and the academic environment is conducive to creative innovation, a spirit of inquiry, and other vital ingredients in the pursuit of educational excellence.

The purpose of the H2 experiment was to get the facts on the relative merits of homogeneous and heterogeneous sectioning in core mathematics. The facts are these: within the limitations of this experiment and the methods of measuring outcomes, no difference in any educationally important outcome is attributable to the difference between homogeneous and heterogeneous sectioning.

In situations in which an experimental comparison is conducted to establish a policy where none previously existed, the null result suggests maintenance of the status quo. In the present situation, a policy existed before the experiment: to *require* homogeneous sectioning in core mathematics courses. In this case, the null result does not suggest maintenance of the status quo. No evidence to support the present policy was discovered. The present policy is not consistent with the results of the experiment under the above axiom. Therefore, it is recommended that core math course directors be allowed to choose freely between homogeneous and heterogeneous modes of sectioning.

Appendix A

PLAN FOR THE 132 H2 EXPERIMENT

Experimental Objectives. The overall experimental objective is to collect and analyze data from the Spring 1979 offering of Math 132 that will enable DFMS management to determine whether heterogeneous or homogeneous sectioning is more desirable for applicable portions of the core mathematics program. The experiment is to be designed in such a way that interference in course operations will be minimal. At the same time, the methodology will be as sound as possible and no data that are readily accessible and that could be used to analyze our educational process will be overlooked.

Assignment of Hours. Math 132 will be offered each of the seven periods on an every day basis. Periods 1, 5 and 7 will be split; that is, both homogeneous and heterogeneous sections will be taught these hours. The reasons to split hours are: (1) it is hypothesized that these hours have special hour effects (sleepiness in 1 and 5 and no athletes in 7). This experimental design will allow us to remove this potential effect as well as test for its existence. (2) The split hours will allow assignment of instructors without having instructors cross treatments. Of the remaining four periods, two will be assigned to each treatment in such a way as to achieve aptitude equivalence between treatment groups.

Assignment of Instructors. Instructors will be assigned randomly to balance experience across treatments. Insofar as possible, instructors will not cross treatments.

Data To Be Collected. (1) Normal grade information; (2) cadet attitudinal survey; (3) instructor attitudinal survey; and (4) EI data.

Criteria To Be Used. (1) Group means; (2) group failure rates; (3) strata means; and (4) any other meaningful criteria that can be constructed from the data collected. Specification of criteria will precede any data analysis.

Course Operation. All guidance will come from the course director on how sections are to be treated. Policy will be that normally promulgated during a similar course, and will be made separately for the two treatments when appropriate.

Sectioning Cadets. DFSCS will section all cadets on a homogeneous basis using 131/131X numerical grade as the sectioning criteria. DFMS will resection the split and heterogeneous hours.

Aptitude-Treatment Interaction. After initial sectioning has been accomplished, treatment groups will be investigated and altered as necessary to ensure equal aptitude profiles (as measured by 131/131X grade) for each.

Data Processing Requirements. One man will be designated, upon approval of this plan, to be responsible for all data processing associated with the experiment. Work will start immediately creating files, planning, developing, and testing programs to be used in the experiment. Workload credit (research) in the spring semester will be given to the man charged with the data processing responsibility.

Data Access. All participants in this experiment, as listed in Responsibilities below, will have access to the experimental data for further exploratory work.

Responsibilities.

Coordinator: LtCol J.D. Sherman

Educational Aspects: Dr. Hassett
Maj Thompson
Capt Webster

Statistical Aspects: Dr. Hassett
LtCol Epp[illegible]n
Maj Mitchell
Maj Thompson

Data Processing: Capt Coffin

Appendix B
SUPPORTING ANALYSES

Table B1. M132/M131 Score by Treatment by Period

Treatment	Period	Mean	Std Dev	Freq
Homogeneous		79.1/76.2	10.8/7.6	429
	1	79.2/76.3	11.8/7.4	59
	2	79.7/77.0	10.3/7.3	125
	5	79.3/76.7	10.1/6.8	71
	6	78.7/75.6	11.5/8.2	117
	7	78.4/74.7	10.5/7.7	57
Heterogeneous		78.6/76.3	10.3/7.6	428
	1	80.0/76.5	9.9/6.9	55
	3	79.0/76.6	11.1/7.9	123
	4	78.9/75.6	9.7/7.8	113
	5	78.4/77.1	9.8/7.2	73
	7	76.3/75.5	10.6/7.5	64

Table B2. ANOVA for Period Effects on M131 Score

Source	df	MS	F
Period	6	69.3	0.3
Residual	850	57.2	

Table B3. ANCOVA for Period Effects on M132 Score with M131 Score as Covariate

Source	df	MS	F
Period	6	40.3	0.8
M131	1	54481.7	1138.2*
Residual	849	47.9	

*$p < .001$

Table B4. Cell Statistics for Instructor Experience Analysis

Treatment	Cadet Aptitude Level	Instructor Experience Level	Mean	Std Dev	Freq
Homogeneous	Low	Inexp	69.7	9.4	92
		Exp	68.4	8.7	49
	Medium	Inexp	79.5	7.7	105
		Exp	80.5	6.4	42
	High	Inexp	87.4	5.1	56
		Exp	88.9	6.0	85
Heterogeneous	Low	Inexp	69.6	8.4	85
		Exp	69.6	9.3	54
	Medium	Inexp	78.5	7.6	88
		Exp	78.7	7.4	58
	High	Inexp	87.1	5.7	83
		Exp	87.6	5.9	60

Table B5. ANOVA for Instructor Experience Effects on M132 Score

Source	df	MS	F
Treatment (T)	1	77.6	1.4
Cadet Aptitude (A)	2	23476.5	419.9*
Experience (E)	1	18.1	0.3
T x A	2	54.1	1.0
T x E	1	2.3	0.0
A x E	2	44.2	0.8
T x A x E	2	26.4	.5
Residual	845	4374.6	

*$p < .001$

Appendix C

OPINION SURVEY FORMS

Instructor Assignment Questionnaire (Pretreatment)

TO:

You are scheduled to teach ____ sections of Math 132 daily in the Spring 1979 semester. An experiment to determine the relative merits of homogeneous and heterogeneous sectioning in Math 132 will be conducted in the Spring semester. Half the Math 132 cadet population will be homogeneously sectioned and half heterogeneously sectioned. Results from questions 1,2,3, and 7 will be used to balance the instructor forces assigned to each sectioning treatment. Results from questions 4,5,6, and 8 will be used to satisfy individual preferences insofar as possible. Of course, it may be impossible to satisfy every instructor's preferences. Please complete the questionnaire and return it to my box NLT 1630, 8 Dec.

1. How many times have you previously taught Math 132?   0   1   2   3 or more

2. How recently have you taught Math 132? ______________________

3. If you will perform flying duties during the Spring 1979 semester, indicate what hours would interfere most with your flying schedule. If none, so state. ______________________

4. Indicate the specific periods that are your first preference for the above number of Math 132 class sections.

   Period   1   2   3   4   5   6   7

5. Indicate the specific periods that are your second preference for the above number of Math 132 class sections.

   Period   1   2   3   4   5   6   7

6. Indicate the specific periods that you prefer not to be scheduled for the above number of Math 132 class sections.

   Period   1   2   3   4   5   6   7

7. Which type of sectioning would you prefer for Math 132?

   (a) homogeneous   (b) undecided   (c) heterogeneous

8. In the event that you are scheduled for homogeneous sections, what cadet aptitude levels would you prefer to teach?

A (highest)	B	C	D	E	F	G	H (lowest)

Instructor Opinion Survey (Post-treatment)

Homogeneous Sectioning - aptitude variance within each section is small, Math 131X. Half of Math 132, and Math 133 were homogeneously sectioned.

Heterogeneous Sectioning - aptitude variance within each section is large. Math 131, and half of Math 132 were heterogeneously sectioned.

Homogeneous Sectioning			Heterogeneous Sectioning	
-2	-1	0	+1	+2
strong opinion	moderate opinion	no difference	moderate opinion	strong opinion

Express your (numerical) strength of opinion to the following questions using the above scale. For example, if you strongly believe that homogeneous sectioning is better for cadets of low math aptitude, record a "-2" for the first question.

Which provides better learning conditions for:

1. Cadets of low math aptitude? ______
2. Cadets of intermediate math aptitude? ______
3. Cadets of high math aptitude? ______
4. The majority of cadets? ______

In questions 5 through 9, consider the whole spectrum of homogeneous sections. Do not answer with specific reference to "A" sections of "H" sections.

5. In which type of sectioning is the instructor's job more difficult? ______
6. Which type of sectioning produces more interaction between cadets in class? ______
7. Which type of sectioning produces the higher EI load? ______
8. Which type of sectioning produces the lower overall instructor workload? ______
9. In which type of sectioning would you rather teach core math? ______
10. Which policy would you recommend to the Department Head for core math courses?
    a. Use heterogeneous sectioning only.
    b. Use homogeneous section only.
    c. Let individual course directors choose type of sectioning.
    d. Other. (explain) ______________________________
    ____
    ____
11. Comments about sectioning needing amplification or not covered by survey items. ______________________________

____

____

____

Table C1. Summary of Instructor Opinion Data

	Homogeneous Sectioning Experience Level												Heterogeneous Sectioning Experience Level											
	0				1				2				0			1			2				
Instructor #	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
Item 1	2	-1	-1	1	2	1	2	-1	-2	2	-1	-2	2	-2	-2	1	1	0	-2	-2	1	2	-1
Item 2	0	-1	-1	2	1	0	0	-2	0	2	0	0	1	-2	0	0	0	1	-2	0	-1	1	0
Item 3	-1	-1	-2	-1	-2	-2	-2	-2	-2	2	-2	-2	-2	-2	1	1	-2	2	-2	-2	-2	-1	-2
Item 4	0	-1	-1	1	2	-1	1	-2	0	2	-1	0	1	-2	-1	1	-1	1	-2	0	0	1	-1
Item 5	-1	-2	1	1	-2	2	1	0	1	-1	-1	2	1	1	1	-1	1	1	1	1	2	-2	1
Item 6	0	0	1	-1	2	-1	2	1	1	2	1	-2	0	1	1	1	0	1	-1	0	2	1	0
Item 7	-1	-2	2	1	-2	-1	0	1	-2	0	-1	-2	0	-1	1	-1	0	0	1	0	2	-1	0
Item 8	0	1	1	-1	2	1	0	0	0	0	0	0	-1	1	-1	1	1	-1	-1	-1	-2	1	-1
Item 9	-1	-1	1	1	2	-1	1	-2	-2	2	-1	-2	1	-2	0	1	-1	1	-2	-2	2	2	-1
Item 10	c	b	a	d	d	b	c	b	b	a	c	b	c	b	d	d	a	d	b	b	b	a	c
Item 7 (pre-treatment)			a		a	a	a	c	a	c	a	a				a	a	a	a	a	b	c	a